

### Analyse von qualitativen abhängigen Variablen

Küsters, Ulrich

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

#### Empfohlene Zitierung / Suggested Citation:

Küsters, U. (1991). Analyse von qualitativen abhängigen Variablen. In H. Best, & H. Thome (Hrsg.), *Neue Methoden der Analyse historischer Daten* (S. 9-40). Sankt Katharinen: Scripta Mercaturae Verl. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-338011>

#### Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

#### Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

# Analyse von qualitativen abhängigen Variablen

von Ulrich Küsters

## 1. Einführung und Übersicht

Dieser Artikel beschreibt die wichtigsten Regressionsmodelle zur Analyse von vermuteten Kausalbeziehungen zwischen nicht-metrischen abhängigen Variablen und unabhängigen Variablen beliebigen Meßniveaus. Regressionsmodelle für metrisch skalierte abhängige Variablen sind weitgehend bekannt. Leider erfüllen die meisten sozialwissenschaftlichen Datensätze nicht die notwendigen Annahmen des gewöhnlichen Regressionsmodells, da die abhängige Variable häufig ordinal oder nominal skaliert ist. Ein typisches Beispiel aus dem historischem Bereich ist die Analyse der Determinanten von nationalsozialistischen Aktivitäten von deutschen Lehrern im Dritten Reich in Abhängigkeit von sozioökonomischen Variablen wie Konfessionszugehörigkeit, Geschlecht, Schichtzugehörigkeit, Geburtskohorte etc. Bei einer derartigen Analyse, die etwa bei Jarausch und Arminger (1989) eingehend behandelt wird, ist die abhängige Variable nominalskaliert, da die verschiedenen Ausprägungen der abhängigen Variablen unterschiedliche NS-Organisationen wie NSDAP, SA, SS etc. repräsentieren, die nicht in einer natürlichen Weise geordnet werden können. Die unabhängigen Variablen weisen hingegen verschiedene Meßniveaus auf. Konfessionszugehörigkeit wird üblicherweise als nominalskalierte Variable aufgefaßt, während eine Variable wie die Größe des Herkunftsortes metrisch skaliert ist.

Wie dieses Beispiel zeigt, muß bei der statistischen Analyse nichtmetrischer Variablen beachtet werden, daß der Wertebereich der abhängigen Variablen eingeschränkt ist. Im klassischen Regressionsmodell wird hingegen unterstellt, daß die abhängige Variable alle Werte auf der reellen Zahlenachse annehmen kann. Daher müssen spezielle Modelle zur Analyse von nichtmetrischen Variablen herangezogen werden, die explizit den beschränkten Wertebereich der abhängigen Variablen berücksichtigen.

Im folgenden wird ein knapper Überblick über die wichtigsten Ansätze zur Analyse nichtmetrischer Variablen gegeben. Zunächst werden die Probleme des kleinsten Quadrateschätzers bei der Analyse von nichtmetrischen Variab-

len skizziert. Anschließend werden Konstruktionsprinzipien zum Aufbau nichtmetrischer Regressionsmodelle beschrieben. Diese Konstruktionsprinzipien basieren im wesentlichen auf der Anwendung des Zufallsnutzenmaximierungsprinzips für nominalskalierte Variablen und des Schwellenwertkonzepts für ordinale und zensierte Variablen. Innerhalb beider Modellklassen werden die wichtigsten Vertreter kurz angegeben. Ein weiterer Abschnitt skizziert einige Probleme bei der Spezifikation und der Schätzung nichtmetrischer Variablenmodelle. Dieser Abschnitt enthält auch einige Hinweise auf statistische Programmsysteme. Zum Abschluß wird ein historischer Datensatz von 743 Abgeordneten aus der Frankfurter Nationalversammlung von 1848/49 exemplarisch analysiert, um die Anwendung qualitativer Variablenmodelle in der empirischen Geschichtsforschung zu illustrieren.<sup>1</sup> Methoden zur Analyse von multivariaten qualitativen Variablen sowie zeitreihenanalytische Methoden werden in diesem Übersichtsartikel nicht behandelt. Detaillierte Beschreibungen der verschiedenen Methoden zur Analyse nichtmetrischer Variablen findet der Leser etwa in den Monographien von Amemiya (1985) und Madala (1983).

## 2. Probleme bei der Anwendung des gewöhnlichen Regressionsmodells auf nichtmetrische Variablen

### Die Schätzung des gewöhnlichen Regressionsmodells

$$y_t = x_t \beta + \varepsilon_t, \quad t = 1, \dots, n$$

basiert auf einer Zufallsstichprobe  $\{y_t, x_t\}_{t=1, \dots, n}$  vom Umfang  $n$ , wobei  $y_t$  eine eindimensionale metrisch skalierte abhängige Variable repräsentiert, während  $x_t$  ein  $1 \times k$  Vektor von unabhängigen Variablen ist, dessen Komponenten entweder metrisch oder dichotom (d.h.  $\{0,1\}$ ) skaliert sind.  $t$  ist der Fallindex. Das lineare Regressionsmodell wird häufig in Matrizenschreibweise durch  $y = X\beta + \varepsilon$  dargestellt. In der Regel werden im klassischen

<sup>1</sup> An dieser Stelle möchte ich mich herzlich bei H. Best, Universität Köln, bedanken, der mir freundlicherweise eine Reihe von Datensätzen zur Verfügung stellte. Herrn R. Ponemereu, vom Zentralarchiv für empirische Sozialforschung, Abteilung Zentrum für historische Sozialforschung Köln, danke ich für das Zusammenstellen eines Datensatzes. G. Armingier (Bergische Universität Wuppertal) und H. Thome (Zentrum für Historische Sozialforschung Köln) gaben mir hilfreiche Kommentare zu einer Vorversion dieser Arbeit.

Regressionsmodell folgende Annahmen unterstellt:

1.  $\{\varepsilon_t\}_{t=1, \dots, n}$  ist eine Folge von stochastisch unabhängigen und identisch verteilten Fehlertermen mit Erwartungswert  $E(\varepsilon_t) = 0$  und endlicher Varianz  $V(\varepsilon_t) = \sigma^2 > 0$ .
2. Die Fehlerterme  $\varepsilon_t$  sind von den Regressoren  $x_t$  stochastisch unabhängig.
3. Die Designmatrix  $X$  hat vollen Spaltenrang. Damit werden lineare Abhängigkeiten zwischen den erklärenden Variablen ausgeschlossen.

Unterstellt man zusätzlich zu diesen fundamentalen Annahmen noch einige asymptotische Regularitätsbedingungen (siehe etwa Schmidt 1976), so folgt, daß der gewöhnliche kleinste Quadrateschätzer (KQS)

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

asymptotisch konsistent und effizient innerhalb der Klasse der asymptotisch normalverteilten Schätzer für  $\beta$  ist. In einer endlichen Stichprobe ist der kleinste Quadrateschätzer  $\hat{\beta}$  erwartungstreu für  $\beta$ .

Offensichtlich hat der kleinste Quadrateschätzer wünschenswerte Eigenschaften in Hinblick auf statistische Qualitätskriterien. Dennoch gibt es eine Reihe von Gründen, den kleinsten Quadrateschätzer für Regressionsmodelle mit nichtmetrischen abhängigen Variablen *nicht* zu verwenden. Zur Illustration der Problematik wird lediglich der Spezialfall einer binären abhängigen Variablen  $y_t$  mit den beiden Ausprägungen 0 und 1 betrachtet. Das Kernproblem des KQS liegt in der Tatsache, daß die KQS-Prädiktoren  $\hat{y}_t = x_t \beta$  nur selten einen der beiden logisch möglichen Werte  $\{0, 1\}$  annehmen. Daher benötigt man ein Interpretationsmodell für die Regressionsgleichung

$$y_t = x_t \beta + \varepsilon_t, \quad t = 1, \dots, n$$

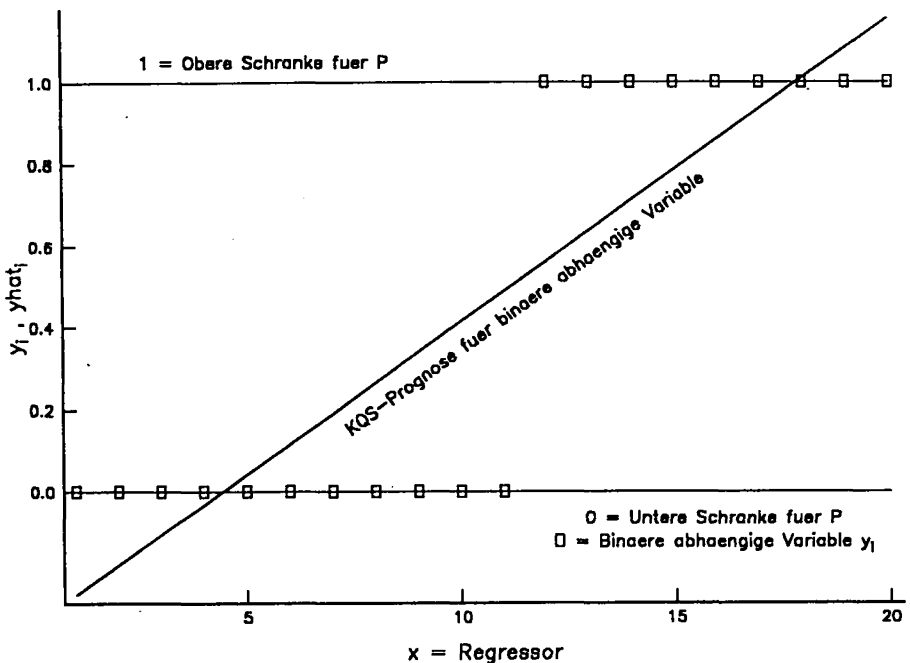
und das Prädiktormodell

$$\hat{y}_t = x_t \hat{\beta}, \quad t = 1, \dots, n$$

Ein einfaches Interpretationsmodell erhält man, indem man die prognostizierten Werte  $\hat{y}_t$  als die geschätzte konditionale Wahrscheinlichkeit der Ausprägung  $y_t = 1$  bei gegebenem Regressor  $x_t$  auffaßt. Dieses Interpreta-

tionsmodell funktioniert jedoch nur dann, wenn der Regressionsvektor  $x_i$  in der Nähe der Mittelwerte der unabhängigen Variablen liegt. Für bestimmte Wertebereiche der unabhängigen Variablen treten jedoch Prädiktoren außerhalb des zulässigen Wertebereichs (0, 1) auf, wie man leicht anhand der Abbildung 1, in der ein einfaches Regressionsmodell mit einem Regressor eingezeichnet ist, erschen kann. Konsequenterweise kann man daher das gewöhnliche Regressionsmodell nicht auf dichotome abhängige Variablen anwenden. Außerdem werden bei dichotomen abhängigen Variablen einige fundamentale Annahmen des linearen Regressionsmodells verletzt. Binäre Variablen sind Bernoulli-verteilt, so daß Heteroskedastizität auftritt, während im klassischen Regressionsmodell homoskedastische Fehlerterme unterstellt werden. Dieses Problem ist jedoch von untergeordneter Wichtigkeit, da man die Heteroskedastizität der Fehler durch verallgemeinerte kleinste Quadrateschätzer berücksichtigen kann. Das Kernproblem ist das Interpretationsmodell. Interpretationsmuster der Form *eine Erhöhung von  $x_i$  um eine Einheit verursacht eine Erhöhung der prognostizierten Wahrscheinlichkeit  $\hat{y}_i$  um  $\hat{\beta}$  Einheiten* sind zwar recht einfach zu handhaben. Jedoch ist völlig unklar, wie prognostizierte negative Wahrscheinlichkeiten und Wahrscheinlichkeiten größer als 1 interpretiert werden können.

Abbildung 1: KQS bei binären Daten



### 3. Konstruktionsprinzipien für Modelle zur Analyse nichtmetrischer Variablen

Fast alle Modelle zur Analyse von nicht-metrischen abhängigen Variablen werden auf der Basis der folgenden Prinzipien konstruiert:

1. Das Prädiktormodell wird grundsätzlich für Wahrscheinlichkeiten und nicht für Ereignisse formuliert.
2. Der Regressionsterm wird weiterhin linear durch  $x_t\beta$  formuliert, so daß man ein einfaches lineares Interpretationsmuster verwenden kann. Damit muß man keine Parameterrestriktionen für  $\beta$  einführen.

Zur Erfüllung dieser Forderungen konzentriert man sich bei der Konstruktion auf Transformationsmodelle. Im folgenden wird der Spezialfall von dichotomen Variablen zuerst behandelt. Anschließend werden Modelle für nominalskalierte, ordinalskalierte und zensierte Variablen dargestellt.

#### 3.1 Modelle zur Analyse von dichotomen Variablen

Zur Sicherung der Interpretation der prognostizierten Werte als Wahrscheinlichkeiten werden binäre Modelle aus zwei Komponenten zusammengesetzt:

1. Ein lineares Prädiktormodell der Form

$$\eta_t = x_t\beta, \quad t = 1, \dots, n$$

wird verwendet, um eine lineare Interpretation der Regressionskoeffizienten  $\beta_t$  bezüglich der latenten Prädiktorvariablen  $\eta_t$  zu ermöglichen.

2. Der lineare Prädiktor  $\eta_t$  wird durch eine nichtlineare Transformation mit einer auf das Intervall  $(0, 1)$  beschränkten Wahrscheinlichkeit für die Ausprägung  $y_t = 1$  verbunden, damit die prognostizierten Werte  $\eta_t$  als Indikatoren für steigende bzw. fallende Wahrscheinlichkeiten interpretiert werden können.

$$P(y_t = 1|x_t) = F(x_t\beta), \quad t = 1, \dots, n$$

Die Funktion  $F: \mathbb{R} \rightarrow (0, 1)$  ist eine monoton steigende stetige Abbildung, die die Wahrscheinlichkeit  $P$  auf das Intervall  $(0, 1)$  beschränkt. Somit ist  $F$  eine Verteilungsfunktion. Üblicherweise wird eine der beiden folgenden Verteilungsfunktionen verwendet:

## 1. Die kumulierte Standardnormalverteilung

$$F(z) = \Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\} dx$$

ergibt das dichotome Probitmodell, das vorwiegend in der Psychometrie und Biometrie angewandt wird.

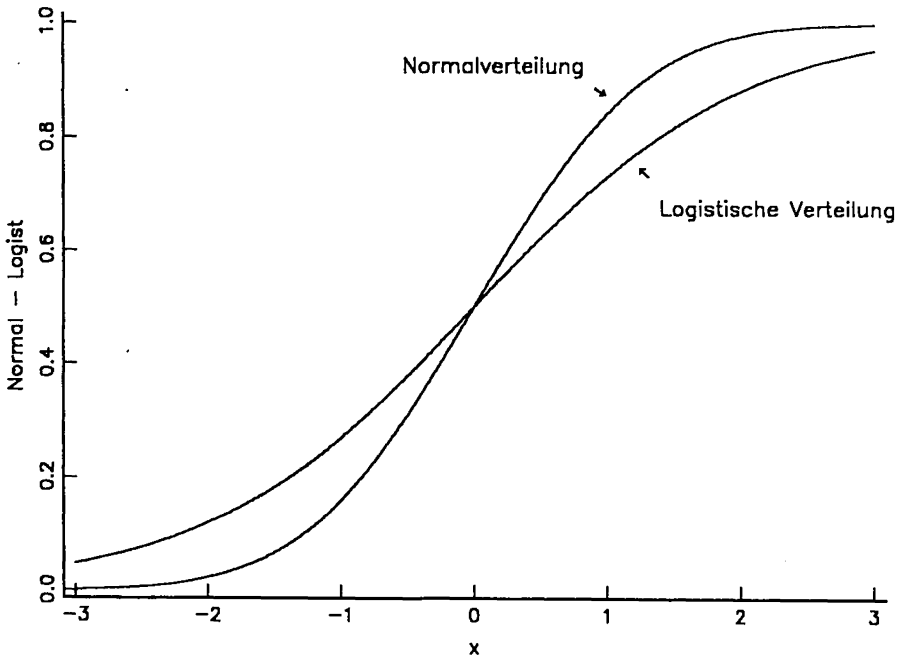
## 2. Die kumulierte logistische Funktion

$$F(z) = L(z) = \frac{1}{1 + \exp(-z)} = \frac{\exp(z)}{1 + \exp(z)}$$

ergibt das dichotome Logitmodell, das vorwiegend in der Soziologie und der Ökonometrie verwendet wird.

Beide Verteilungsfunktionen sind in Abbildung 2 eingetragen und stimmen mit Ausnahme der Ränder nahezu überein, sofern man die Dichte auf identische Varianzen skaliert. Daher erhält man bei praktischen Analysen bis auf einen Skalenfaktor nahezu identische Ergebnisse, so daß die Selektion zwischen beiden Verfahren im wesentlichen eine Frage der verfügbaren Software ist.

Abbildung 2: Normalverteilung – Logistische Verteilung



Bisher wurde davon ausgegangen, daß die unabhängigen Variablen vollständig metrisch skaliert sind. Daher wird in diesem Abschnitt skizziert, welche Modifikationen eingeführt werden müssen, wenn zusätzlich oder ausschließlich nominalskalierte Regressoren auftreten. Der einfachste Ansatz zur Berücksichtigung von nominalskalierten Regressoren ist die Verwendung von Designmatrizen, wie sie seit langem bei der metrischen Varianzanalyse verwendet werden. Zur Analyse einer metrischen abhängigen Variablen  $y_t$  in Abhängigkeit von einer ungeordneten qualitativen Variablen  $x_{t1}$  mit  $C$  Ausprägungen wird die Regressionsgleichung durch Mittelwerte strukturiert:

$$y_t = \mu + x_{t11}\alpha_1 + x_{t12}\alpha_2 + \dots + x_{t1C}\alpha_C + \varepsilon_t, \quad t = 1, \dots, n$$

$x_{tli}$  ist eine Dummy-Variable (Indikatorvariable) mit

$$x_{tli} = \begin{cases} 1 & \text{wenn die abhängige Variable } x_{t1} \text{ bei Fall } t \\ 0 & \text{sonst} \end{cases} \quad \text{Ausprägung } i \text{ annimmt}$$

Damit repräsentiert der zusammengesetzte Term  $\mu + \alpha_i$  den Mittelwert der abhängigen Variablen  $y_t$  unter der Bedingung, daß die unabhängige Variable  $x_{t1}$  die Ausprägung  $i$  annimmt. Allerdings enthält das Modell  $C + 1$  Parameter, obwohl aus den Daten lediglich  $C$  konditionale Mittelwerte geschätzt werden können. Daher muß eine zusätzliche Parameterrestriktion eingeführt werden.

Meistens wird einer der beiden folgenden Ansätze<sup>2</sup> verwendet: Am einfachsten ist die Restriktion  $\sum_{i=1, \dots, C} \alpha_i = 0$ , so daß  $\mu$  dem Mittelwert von  $y_t$  in der gesamten Stichprobe entspricht. Dieser erste Ansatz wird in der Literatur als zentrierte Reparametrisierung bezeichnet. Leider treten bei diesem Ansatz statistische Probleme auf, wenn Stichprobennullen oder strukturelle Nullen (siehe unten) auftreten. Der zweite Ansatz, der als Eckpunktparametrisierung bezeichnet wird, restringiert den Parameter der ersten Kategorie auf  $\alpha_1 = 0$ . Unter dieser Reparametrisierung korrespondiert der geschätzte Mittelwert  $\mu$  zum Mittelwert von  $y_t$  innerhalb der Stichprobenpartition mit  $x_{t1} = 1$ , während die geschätzten Parameter  $\alpha_i$  als Mittelwertverschiebungen gegenüber der ersten Kategorie interpretiert werden können. Beide Reparametrisierungsformen können einfach in das lineare Regressionsmodell eingebettet werden. Die Eckpunktrepametrisierung wird in Regressionsschreibweise durch folgende Designmatrix repräsentiert:

<sup>2</sup> Komplexere Designmatrizen findet der Leser in Bock (1975).



$$\begin{pmatrix} \text{Prädiktor für Kategorie 1} \\ \text{Prädiktor für Kategorie 2} \\ \text{Prädiktor für Kategorie 3} \\ \text{Prädiktor für Kategorie 4} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} \mu \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix} = \begin{pmatrix} \mu \\ \mu + \alpha_2 \\ \mu + \alpha_3 \\ \mu + \alpha_4 \end{pmatrix}$$

Beide Ansätze können direkt auf den Fall mit mehreren unabhängigen Variablen verallgemeinert werden. Angenommen, zwei kategoriale Variablen  $x_{i1}$  und  $x_{i2}$  sollen zur Erklärung der Subgruppenmittelwerte einer metrischen abhängigen Variablen  $y_i$  herangezogen werden. Ein adäquates Prädiktormodell ist durch folgende Gleichung gegeben:

$$\hat{y}_i = \mu + \sum_{i=1}^{C_1} (\alpha_i x_{i1i}) + \sum_{j=1}^{C_2} (\gamma_j x_{i2j}) + \sum_{i=1}^{C_1} \sum_{j=1}^{C_2} (\alpha\gamma_{ij} x_{i1i} x_{i2j})$$

Die Parameter  $\alpha_i$  und  $\gamma_j$  repräsentieren die direkten Effekte der Dummy-Variablen  $x_{i1i}$  bzw.  $x_{i2j}$  auf  $y_i$ , während die sogenannten Interaktionseffekte  $\alpha\gamma_{ij}$  die zusätzlichen kombinierten Effekte beider Variablen messen, die nicht durch die direkten (additiven) Effekte berücksichtigt werden. Auch im multivariaten Fall müssen Parameterrestriktionen eingeführt werden, um die Schätzbarkeit der Parameter zu sichern. Bei einer Eckpunktparametrisierung wird üblicherweise die Identifikationsrestriktion  $\alpha_i = \gamma_j = \alpha\gamma_{ij} = 0$  für alle Indizes mit  $i = 1$  oder  $j = 1$  verwendet. Somit kann man die varianzanalytische Designformulierung auch direkt für binäre Modelle verwenden, da die Struktur der linearen Prädiktorform  $\eta_i = x_i\beta$  (bzw.  $\eta = X\beta$ ) in binären Modellen mit der Prädiktorform der Varianzanalyse übereinstimmt.

### 3.2 Modelle zur Analyse von nominalskalierten abhängigen Variablen

Konzeptionell basieren die meisten Regressionsmodelle zur Analyse ungeordneter kategorialer Variablen auf einer Anwendung des Zufallsnutzenmaximierungsprinzips, bei dem jeder Ausprägung (Kategorie)  $i \in \{1, \dots, C\}$  der abhängigen Variablen ein Zufallsnutzen<sup>3</sup>  $U_{ii}$  zugeordnet wird, der durch

<sup>3</sup> Das Zufallsnutzenmaximierungsprinzip wird in dieser Arbeit lediglich als mathematisches Konzept benutzt. Zur Modellierung von Regressionsmodellen mit qualitativen abhängigen Variablen kann man den Term  $U_{ii}$  genauso gut als eine latente Variable auffassen, die die relative Neigung zu Kategorie  $y_i = i$  bei gegebenem Regressor  $x_i$  charakterisiert. Allerdings implizieren statistische Zufallsmaximierungsmodelle eine Reihe von nutzentheoretischen Eigenschaften, die in der Wahltheorie (McFadden 1981) ausführlich behandelt werden.

eine additive Kombination einer linearen Funktion  $x_i \beta_i$  mit einem Fehlerterm  $\varepsilon_{it}$  parametrisiert wird.

$$U_{it} = x_i \beta_i + \varepsilon_{it}, \quad i = 1, \dots, C; \quad t = 1, \dots, n$$

Die Parameter  $\beta_i$  und die Fehlerterme  $\varepsilon_{it}$  stellen kategorienspezifische Größen dar. Nach dem Zufallsnutzenmaximierungsprinzip ist die konditionale Selektionswahrscheinlichkeit der Kategorie  $y_t = i$  bei gegebenem  $x_t$  durch den Ausdruck

$$P(y_t = i | x_t) = P(U_{it} > U_{jt} \text{ für alle } j = 1, \dots, C \text{ mit } j \neq i)$$

definiert. Man beachte, daß die Zufallsnutzenterme  $U_{it}$  beliebige Werte auf der reellen Zahlenachse annehmen können. Dennoch sind alle Selektionswahrscheinlichkeiten auf das Intervall  $(0, 1)$  beschränkt. In Analogie zu binären Logit- und Probitmodellen werden verschiedene Modelle durch unterschiedliche Verteilungsannahmen für  $\varepsilon_{it}$  generiert. Zwei typische Vertreter dieser Modellklasse sind:

1. Spezifiziert man die Verteilung des Fehlertermvektors  $\varepsilon_t = (\varepsilon_{t1}, \dots, \varepsilon_{tC})^T$  durch eine  $C$ -dimensionale Normalverteilung mit Erwartungswert 0 und Varianz-Kovarianzmatrix  $\Sigma$ , so erhält man das multinomiale Probitmodell (Daganzo 1979).
2. Das multinomiale Logitmodell (McFadden 1974, Bock 1975) erhält man hingegen unter der Annahme, daß die einzelnen Komponenten  $\varepsilon_{it}$  des Fehlertermvektors  $\varepsilon_t$  stochastisch unabhängig und identisch verteilt sind und einer Extremwertverteilung vom Typ I mit der Verteilungsfunktion

$$F(z) = \exp(-\exp(-z))$$

folgen.

Der wesentliche Vorteil des multinomialen Probitmodells gegenüber dem multinomialen Logitmodell besteht in der weitaus größeren Flexibilität, da Ähnlichkeiten zwischen den verschiedenen Ausprägungen durch positive Korrelationsparameter in  $\Sigma$  berücksichtigt werden können. Ein wesentlicher Nachteil ist hingegen die Tatsache, daß multinomiale Probitmodelle numerisch nur sehr aufwendig für abhängige Variablen mit mehr als 4 Kategorien (= Ausprägungen) geschätzt werden können, da bei der Schätzung eines Modells mit  $C$  Kategorien  $C - 1$  dimensionale Normalverteilungsintegrale für jeden Fall berechnet werden müssen. Daher wird das multinomiale Logitmodell üblicherweise vorgezogen, da die Schätzung bedeutend einfacher ist. Dies ist darauf zurückzuführen, daß die Selektionswahrscheinlichkeiten im multinomialen Logitmodell explizit durch den Ausdruck

$$P(y_i = i | x_i) = \frac{\exp(x_i \beta_i)}{1 + \sum_{l=2}^C \exp(x_i \beta_l)}, \quad i = 2, \dots, C$$

gegeben sind, während eine geschlossene Formel für die Selektionswahrscheinlichkeiten des multinomialen Probitmodells nicht zur Verfügung steht. Zur Sicherung der Parameterschätzbarkeit wird üblicherweise die Restriktion  $\beta_1 = 0$  verwendet. Das multinomiale Logitmodell gehört zur Klasse der Exponentialfamilienmodelle. Somit ist die Log-Likelihood-Funktion konkav, so daß bei der numerischen Bestimmung des Maximum-Likelihood-Schätzers nur relativ wenige Probleme auftreten.

Individualdaten lassen sich ohne Informationsverlust zu Kontingenztabellen aggregieren, wenn alle Variablen nominalskaliert sind. Zur Analyse von Kontingenztabellen werden häufig loglineare Modelle (Bishop, Fienberg und Holland 1975, Jarausch, Arminger und Thaller 1985) herangezogen, mit denen die absoluten Häufigkeiten der einzelnen Zellen modelliert werden. Damit gehören loglineare Modelle in ihrer allgemeinsten Form zur Klasse der symmetrischen Modelle, da keine Variable explizit als abhängig ausgewiesen wird. Durch geeignete Parametrisierungsansätze kann man jedoch auch die Regressionsparameter  $\beta_i$  eines multinomialen Logitmodells indirekt durch die Schätzung eines loglinearen Modells bestimmen. Als Beispiel einer derartigen Parametrisierung wird eine dreidimensionale Kontingenztafel mit drei Variablen  $\{Z_1, Z_2, Z_3\}$  und den Zellenbesetzungen  $n_{ijk}$  betrachtet. Die Parametrisierung des loglinearen Modells zur exakten Reproduktion der gemeinsamen, dreidimensionalen Verteilung der Variablen  $\{Z_1, Z_2, Z_3\}$  lautet

$$\ln(n_{ijk}) = \alpha_i + \gamma_j + \delta_k + \alpha\gamma_{ij} + \alpha\delta_{ik} + \gamma\delta_{jk} + \alpha\gamma\delta_{ijk}$$

Dabei sind  $i = 1, \dots, C_{Z_1}$ ;  $j = 1, \dots, C_{Z_2}$  und  $k = 1, \dots, C_{Z_3}$  die Laufindizes der Ausprägungen der drei Variablen. In Analogie zur Varianzanalyse müssen die Effektparameter restringiert werden, um die Schätzbarkeit der Parameter zu sichern. Bei einer Eckpunktreparametrisierung werden üblicherweise die zu den ersten Kategorien korrespondierenden Effekte  $\alpha_1, \gamma_1, \delta_1, \alpha\gamma_{ij}, \alpha\delta_{ik}, \gamma\delta_{jk}$  und  $\alpha\gamma\delta_{ijk}$  für  $i, j, k = 1$  auf Null restringiert. Ein Modell, mit dem die gemeinsame Verteilung aller Variablen exakt reproduziert werden kann, wird als saturiert (gesättigt) bezeichnet. Bei einer statistischen Modellierung werden i.d.R. einige Parameter auf Null restringiert, so daß das Modell die gemeinsame Verteilung nicht exakt reproduziert. Zur Erklärung der dreidimensionalen, gemeinsamen Verteilung der drei Variablen  $\{Z_1, Z_2, Z_3\}$

durch die drei bivariaten Verteilungen  $\{Z1, Z2\}$ ,  $\{Z1, Z3\}$  und  $\{Z2, Z3\}$  verwendet man das reduzierte Modell

$$\ln(n_{ijk}) = \alpha_i + \gamma_j + \delta_k + \alpha\gamma_{ij} + \alpha\delta_{ik} + \gamma\delta_{jk} + \varepsilon,$$

in dem alle Interaktionseffekte dritter Ordnung  $\alpha\gamma\delta_{ijk}$  auf Null restringiert werden. Möchte man die dreidimensionale Verteilung durch die univariaten marginalen Verteilungen modellieren (Unabhängigkeitsmodell), so werden nur die Haupteffekte  $\alpha_i$ ,  $\gamma_j$  und  $\delta_k$  zur Erklärung der gemeinsamen Häufigkeiten  $n_{ijk}$  geschätzt. Die bisher angegebenen Parametrisierungsformen von loglinearen Modellen weisen keiner Variablen eine besondere Bedeutung zu, so daß man diese Modelle in Analogie zur Korrelationsanalyse als reine Zusammenhangsmodelle interpretieren kann. Möchte man hingegen die Variable Z1 durch die beiden Variablen  $\{Z2, Z3\}$  erklären, so kann man das loglineare Modell mit Haupteffekten und Interaktionseffekten erster Ordnung in folgender Weise als multinomiales Logitmodell mit den Regressoren  $\{Z2, Z3\}$  auffassen:

$$P(Z1 = i | Z2 = j, Z3 = k) = \frac{n_{ijk}}{C_{Z1}} = \frac{\sum_{l=1}^{C_{Z1}} n_{ljk}}{\sum_{l=1}^{C_{Z1}} \exp(\alpha_l + \gamma_j + \delta_k + \alpha\gamma_{lj} + \alpha\delta_{lk} + \gamma\delta_{jk})} = \frac{\exp(\alpha_i + \gamma_j + \delta_k + \alpha\gamma_{ij} + \alpha\delta_{ik} + \gamma\delta_{jk})}{\sum_{l=1}^{C_{Z1}} \exp(\alpha_l + \gamma_j + \delta_k + \alpha\gamma_{lj} + \alpha\delta_{lk} + \gamma\delta_{jk})} = \frac{\exp(\alpha_i + \alpha\gamma_{ij} + \alpha\delta_{ik})}{\sum_{l=1}^{C_{Z1}} \exp(\alpha_l + \alpha\gamma_{lj} + \alpha\delta_{lk})}$$

Damit kann man die Parameter  $\alpha_i$ ,  $\alpha\gamma_{ij}$  und  $\alpha\delta_{ik}$  als Regressionsparameter  $\beta_i = (\alpha_i, \alpha\gamma_{ij}, \alpha\delta_{ik})^T$  eines multinomialen Logitmodells auffassen, in dem die konditionalen Selektionswahrscheinlichkeiten der abhängigen Variablen Z1 durch die beiden unabhängigen kategorialen Variablen Z2 und Z3 erklärt werden. Somit können Regressionsparameter von multinomialen Logitmodellen grundsätzlich durch die Schätzung eines äquivalenten loglinearen

Modells bestimmt werden, sofern alle unabhängigen Variablen nominalskaliert sind<sup>4</sup>.

Wendet man das multinomiale Probitmodell und das multinomiale Logitmodell auf eine binäre abhängige Variable an, so erhält man die korrespondierenden binären Probit- bzw. Logitmodelle. Man beachte jedoch, daß sich beide Modelle im Gegensatz zum binären Fall *erheblich* unterscheiden, wenn die abhängige Variable mehr als zwei Kategorien besitzt. Heuristisch kann dieser Unterschied an folgendem Musterbeispiel erklärt werden: Angenommen, man möchte die Konsumentenwahl zwischen drei verschiedenen Transportmodi (z.B. Bus, Bahn und Auto) zum Arbeitsplatz durch eine Reihe von exogenen Regressoren erklären. Bus und Bahn gehören zur Klasse der öffentlichen Verkehrsmittel, die in der Regel durch gemeinsame oder ähnliche Attribute charakterisierbar sind, die nicht durch unabhängige Variablen gemessen werden. Der Transportkomfort ist ein klassisches Beispiel. Derartige Ähnlichkeiten zwischen Teilmengen von Alternativen stellen unbeobachtbare Heterogenitätsfaktoren dar, die in einem multinomialen Probitmodell relativ einfach durch eine positive Fehlertermkorrelation in  $\Sigma$  zwischen den Ausprägungen Bus und Bahn modelliert werden können. Innerhalb eines multinomialen Logitmodells kann ein derartiger Heterogenitätseffekt jedoch nicht berücksichtigt werden, da grundsätzlich unterstellt wird, daß die Fehlerterme stochastisch unabhängig sind. Aufgrund der erheblichen numerischen Probleme bei der Schätzung des multinomialen Probitmodells wurden in der Literatur eine Reihe von Modifikationen des multinomialen Logitmodells vorgeschlagen, die zumindest eine partielle Berücksichtigung von Ähnlichkeiten zwischen den verschiedenen Kategorien erlauben. Die wichtigsten Vertreter dieser Modellklasse sind die genesteten (= geschachtelten) Logitmodelle und die Tree-Extreme-Value-Modelle von McFadden (1981).

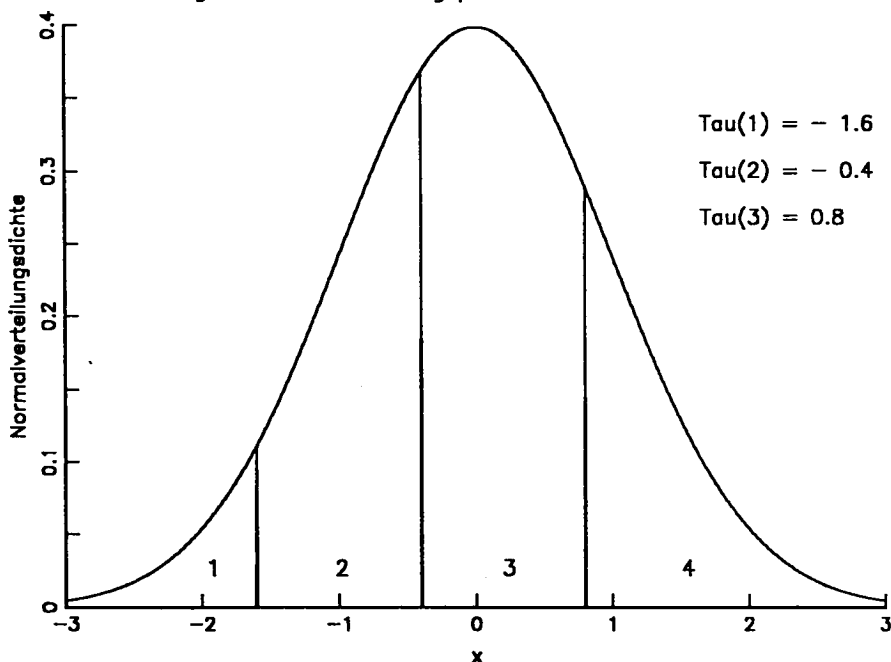
### 3.3 Modelle zur Analyse von ordinalskalierten Variablen

Zufallsnutzenmaximierungsmodelle zur Erklärung der Variation von ungeordneten qualitativen Variablen basieren auf der Tatsache, daß die einzelnen Ausprägungen der abhängigen Variablen *nicht* auf einer Rangskala angeordnet werden können. Bei praktischen Analysen treten jedoch relativ häufig abhängige Variablen auf, die in einer natürlichen Rangfolge angeord-

<sup>4</sup> Der Satz von Birch (siehe Bishop, Fienberg und Holland 1975) liefert eine mathematische Grundlage für die Äquivalenz beider Ansätze.

net werden können. Ein typisches Beispiel ist die Messung von Einstellungen auf einer Likert-Skala. Eine spezielle Modellklasse, die eine derartige zusätzliche Ranginformation ausnutzt, verwendet das Schwellenwertkonzept. Schwellenwertmodelle basieren auf der grundlegenden Annahme, daß eine unbeobachtbare (latente) Variable  $\eta_i$  existiert, die alle Werte auf der reellen Zahlenachse annehmen kann. Diese latente Zufallsvariable wird durch ein lineares Regressionsmodell parametrisiert. Steigende Werte von  $\eta_i$  können als verstärkte relative Neigung von Kategorie  $y_i = i$  im Verhältnis zu einer Referenzkategorie (etwa  $y_i = 1$  oder  $y_i = C$ ) interpretiert werden. Da die Ausprägungen geordnet sind, kann man lediglich die niedrigste oder die höchste Kategorie als Referenzkategorie wählen. Zur Verbindung der latenten Variablen  $\eta_i$  mit den diskreten Ausprägungen  $y_i = 1, \dots, C$  wird die reelle Zahlenachse  $R$  durch  $C - 1$  geordnete Schwellenwerte  $\tau_1 < \tau_2 < \dots < \tau_{(C-1)}$  partitioniert, wie man aus Abbildung 3 ersehen kann.

Abbildung 3: Normalverteilung partitioniert durch Schwellenwerte



Die abhängige Variable  $y_t$  nimmt genau dann den Wert  $i$  an, wenn die latente Variable  $\eta_t$  in das halboffene Intervall  $(\tau_{i-1}, \tau_i]$  fällt. Bei festem Regressor  $x_t$  wird die stochastische Variation von  $\eta_t$  durch die stochastische Variation des Fehlerterms  $\varepsilon_t$  verursacht. Formal können Schwellenwertmodelle durch folgende Gleichungen repräsentiert werden:

$$\eta_t = x_t \beta + \varepsilon_t, \quad t = 1, \dots, n$$

und

$$y_t = i \Leftrightarrow \eta_t \in (\tau_{i-1}, \tau_i], \quad i = 1, 2, \dots, C, \quad t = 1, \dots, n.$$

Dabei wird angenommen, daß die unbekannten und damit zu schätzenden Schwellenwerte durch  $-\infty = \tau_0 < \tau_1 < \dots < \tau_{C-1} < \tau_C = +\infty$  geordnet sind.

Unterschiedliche Modelle werden wiederum durch verschiedene Fehlertermannahmen generiert. Unterstellt man, daß die Fehlerterme  $\varepsilon_t$  stochastisch unabhängig und identisch standard-normalverteilt sind, so erhält man das ordinale Probitmodell. Dies entspricht der Verteilung in Abbildung 3. Ist der Fehlerterm  $\varepsilon_t$  logistisch verteilt, so erhält man das ordinale Logitmodell (Bock 1975). Ähnliche Modelle können durch andere Verteilungsannahmen für  $\varepsilon_t$  erzeugt werden (McCullagh 1980). Übrigens unterstellen alle erwähnten Modelle, daß die Folge der Fehlerterme  $\{\varepsilon_t\}_{t=1, \dots, n}$  stochastisch unabhängig ist.

Eine natürliche Rangfolge zwischen den Kategorien der abhängigen Variablen reicht im allgemeinen nicht zur Anwendung eines ordinalen Schwellenwertmodells aus. Diese Einschränkung ergibt sich aus der einfachen Überlegung, daß die Effekte der unabhängigen Variablen einen U-Effekt in dem Sinn bewirken können, daß sowohl extrem hohe als auch extrem niedrige Werte einer unabhängigen Variablen zu hohen Neigungen bezüglich einer Kategorie der abhängigen Variablen führen können, während mittlere Werte der unabhängigen Variablen zu niedrigen Neigungen führen. In einer derartigen Situation stehen zwei Analysealternativen zur Verfügung. Erstens kann man in Analogie zur nichtlinearen metrischen Regression ein in  $x_t$  nichtlineares Prädiktormodell für  $\eta_t$  formulieren. Andererseits kann man aber auch einfach die natürliche Rangfolge zwischen den Kategorien der abhängigen Variablen ignorieren und ein Modell zur Analyse von nominalskalierten abhängigen Variablen verwenden. Das stereotype Modell von Anderson (1984) basiert auf einer Modifikation des multinomialen Logitmodells, in dem spezielle Parametrisierungen zur Berücksichtigung von Parameteranordnungen

verwendet werden.

Gruppierte metrische abhängige Variablen, bei denen lediglich die Klassenzugehörigkeit zu einem Intervall bekannt ist, werden in der Regel analog zu ordinalen Variablen behandelt, indem die Schwellenwerte auf die festen Intervallgrenzen restringiert werden. Zusätzlich muß die Varianz des Fehlertermvektors freigesetzt werden (Stewart 1983).

### 3.4 Modelle zur Analyse von zensierten Variablen

Bei einer zensierten Variablen  $y_i$  kann der exakte Wert der abhängigen Variablen lediglich für eine Teilmenge der Stichprobenelemente  $t = 1, \dots, n$  beobachtet werden. Beobachtungen innerhalb der Komplementärmenge werden hingegen nur durch eine Indikatorvariable (etwa durch 0 aus dem Indikatorset  $\{0, 1\}$ ) repräsentiert, ohne daß der exakte Wert der abhängigen Variablen bekannt ist. Derartige Variablen treten oft bei Untersuchungen zum Verhalten von Konsumenten beim Kauf von dauerhaften Gebrauchsgütern auf, bei denen die abhängige Variable den Kaufbetrag eines bestimmten Gebrauchsgutes repräsentiert. Eine ganze Reihe von dauerhaften Konsumgütern (etwa Autos und Waschmaschinen) werden nicht von allen Konsumenten innerhalb einer fixen Haushaltsperiode (etwa ein Jahr) gekauft, sondern nur in Abständen von mehreren Jahren erneuert. Damit repräsentiert die Ausprägung 0 der Indikatorvariablen das Unterlassen eines Kaufaktes. Wird ein Kauf getätigt, so sind die Beträge grundsätzlich positiv. Wendet man den gewöhnlichen KQS auf zensierte Daten an, so erhält man inkonsistente und verzerrte Schätzungen, wie man anhand der Abbildung 4 ersehen kann.

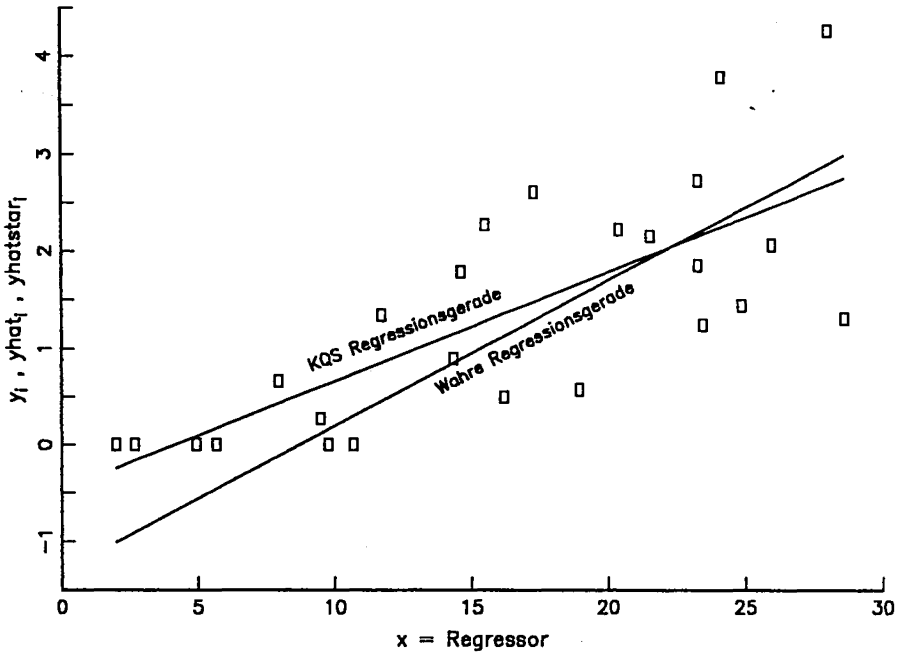
Offensichtlich ist der gewöhnliche KQS nicht erwartungstreu und konsistent, da die auf Null zensierten Werte den Steigungsparameter  $\beta_1$  nach unten verzerren. Verwendet man lediglich diejenigen Stichprobenelemente mit positiven  $y_i$  Beobachtungen, so wird dieser Fehler zwar korrigiert. Allerdings ist diese Vorgehensweise ineffizient, da man auf die in den zensierten Fällen enthaltene Information verzichtet.

Ein adäquates Modell zur Behandlung von zensierten Variablen<sup>5</sup> ist das Tobit-Modell von Tobin (1959), das eine Mischung des gewöhnlichen Regressionsmodells mit einem binären Probitmodell darstellt. Bei der Modellierung von Tobitmodellen wird folgende Grundstruktur (Amemiya 1973) unterstellt:

<sup>5</sup> Man beachte, daß der hier verwendete Zensierungsbegriff nur sehr wenig mit den Zensierungsbegriffen in der "Failure Time Series" Literatur zu tun hat.



Abbildung 4: KQS bei zensierten Daten



$$\eta_t = x_t \beta + \varepsilon_t, \quad t = 1, \dots, n$$

$\varepsilon_t$  ist eine Folge von stochastisch unabhängigen und identisch normalverteilten Zufallsvariablen mit Erwartungswert 0 und unrestringierter Varianz  $\sigma^2 > 0$ . Die Meßrelation wird durch folgende Gleichung definiert:

$$y_t = \begin{cases} \eta_t & \text{wenn } \eta_t \text{ einen bekannten Schwellenwert } \tau \text{ überschreitet} \\ \tau & \text{sonst} \end{cases}$$

Im Kontext von Untersuchungen zur Konsumentennachfrage kann die partiell beobachtbare latente Variable  $\eta_t$  als marginale Kaufneigung interpretiert werden. Das oben skizzierte Tobit-Modell ist lediglich ein exemplarisches Beispiel einer weiten Klasse von Tobit-Modellen, die eingehend in Amemiya (1985) dargestellt wird. Eine wichtige Verallgemeinerung parametrisiert den festen Schwellenwert  $\tau$  durch eine stochastische oder deterministische Regressionsgleichung.

## 4. Schätzung und Modellspezifikation

Mit Ausnahme des multinomialen Probitmodells ist die numerische Schätzung der Parameter in qualitativen Variablenmodellen recht einfach, da die Loglikelihood-Funktionen in der Regel ein eindeutiges Maximum besitzen, so daß man relativ einfache numerische Optimierungsverfahren wie Newton-Raphson und Fisher's Scoring (Thisted 1988) verwenden kann. Modellspezifische Probleme findet der Leser in den Monographien von Maddala (1983) und Amemiya (1985). Daher beschränkt sich dieser Abschnitt auf eine knappe Diskussion von speziellen Datensatzkonstellationen, Modellspezifikationsprozeduren und auf eine selektive Angabe problemadäquater Softwarepakete.

### 4.1 Strukturelle Nullen und Stichprobennullen

Bei Datensätzen mit qualitativen Variablen tritt sehr häufig das Problem auf, daß eine Teilmenge der Ausprägungen innerhalb von Teilstichproben, die sich durch gemeinsame Regressoren charakterisieren lassen, nicht besetzt sind. Die Ursachen für derartige Nullbesetzungen lassen sich durch folgende Unterscheidung klassifizieren:

1. Innerhalb einer Teilstichprobe ist eine bestimmte Ausprägung der abhängigen Variablen logisch unmöglich. Ein einfaches Beispiel dieses Falles, der in der Literatur als strukturelle Null bezeichnet wird, ist die Tatsache, daß Frauen in nationalsozialistischen Organisationen wie der SS nicht zugelassen wurden. Daher kann die Variable Geschlechtszugehörigkeit nicht zur Erklärung der SS-Mitgliedschaft herangezogen werden.
2. Weitaus häufiger tritt jedoch der Fall auf, daß bestimmte (i.d.R. seltene) Kombinationen von Ausprägungen der abhängigen und der unabhängigen Variablen aufgrund unzureichendem Stichprobenumfangs nicht besetzt sind. Dieser Fall wird als Stichprobennull bezeichnet.

Nullbesetzungen können unter bestimmten Parametrisierungen zu einer Nichtexistenz des Maximum-Likelihood-Schätzers führen, obwohl sich dieses Problem bei Stichprobennullen durch eine selektive Nacherhebung oder durch eine Vergrößerung des Stichprobenumfangs eliminieren läßt. Das resultierende Schätzproblem wird im folgenden an einem einfachen Beispiel einer Vierfeldertafel illustriert, die theoretisch mit Hilfe eines binären Logitmodells analysiert wird:

Geschlecht / Organisation	SS-Mitglieder	keine SS-Mitglieder
Männlich	$n_{11}$	$n_{12}$
Weibliche	0	$n_{22}$

Man beachte, daß die relative Häufigkeit einer SS-Mitgliedschaft innerhalb der Gruppe der Männer durch  $n_{11}/(n_{11}+n_{12})$  gegeben ist, während die relative Häufigkeit einer SS-Mitgliedschaft von Frauen *a priori* auf Null restringiert ist. Eine Parametrisierung zur Schätzung der Effekte von Geschlecht auf Mitgliedschaft mit einem binären Logitmodell ergibt folgende Modellstruktur:

$$P(\text{SS Mitglied} | \text{Männlich}) = \frac{\exp(\mu)}{1 + \exp(\mu)}$$

$$P(\text{SS Mitglied} | \text{Weiblich}) = \frac{\exp(\mu + \alpha)}{1 + \exp(\mu + \alpha)}$$

$$P(\text{Kein SS Mitglied} | \text{Männlich}) = 1 - P(\text{SS Mitglied} | \text{Männlich})$$

$$P(\text{Kein SS Mitglied} | \text{Weiblich}) = 1 - P(\text{SS Mitglied} | \text{Weiblich})$$

Offensichtlich können die Parameter  $\mu$  und  $\alpha$  exakt an die Daten angepaßt werden, da in der allgemeinen Formulierung des Modells nur zwei freie Wahrscheinlichkeiten, nämlich  $P(\text{SS Mitglied} | \text{Männlich})$  und  $P(\text{SS Mitglied} | \text{Weiblich})$  auftreten. Daher kann der Maximum-Likelihood-Schätzer von  $\mu$  analytisch durch den Ausdruck

$$\hat{\mu} = \ln\left(\frac{n_{11}}{n_{12}}\right)$$

bestimmt werden. Ist  $\mu$  gegeben, so kann man die absolute Häufigkeit von  $n_{21} = 0$  nur dann exakt an

$$P(\text{SS Mitglied} | \text{Weiblich}) = \frac{\exp(\mu + \alpha)}{1 + \exp(\mu + \alpha)}$$

anpassen, wenn man den Parameterschätzer für  $\alpha$  gegen  $-\infty$  konvergieren läßt. Somit existiert in diesem Modell kein endlicher Maximum-Likelihood-Schätzer für  $(\mu, \alpha)$ . Verwendet man zur numerischen Bestimmung des Maximum-Likelihood-Schätzers ein iteratives Optimierungsverfahren wie das Newton-Raphson-Verfahren, so kann man diesen Effekt sehr leicht an einer

Divergenz einer Parametersubgruppe entdecken. Falls ein infinites Maximum-Likelihood-Schätzer auftritt, so muß man die divergierenden Parameter durch eine Reparametrisierung aus der Modellspezifikation eliminieren. Leider sind eine Reihe von Softwarepaketen nicht in der Lage, eine Parameterdivergenz zu diagnostizieren. Dies liegt zum Teil daran, daß einige Programmsysteme ungeeignete Schätzverfahren wie den einstufigen GSK-Schätzer verwenden. Anderere Systeme benutzen hingegen ungeeignete Abbruchkriterien wie Gradientenabstände zu Null. Ein Anwender muß daher diese Probleme bei der Auswahl und der Anwendung von Software berücksichtigen. Erwähnenswert ist die Tatsache, daß eine Nichtexistenz des ML-Schätzers grundsätzlich an eine bestimmte Parametrisierung gebunden ist. Restrangiert man im obigen Beispiel den Regressionsparameter  $\alpha$  auf Null, so existiert der eindimensionale ML-Schätzer  $\mu$ , so daß man unter gewissen Einschränkungen das Existenzproblem des ML-Schätzers durch eine eingeschränkte Parametrisierung umgehen kann. Allerdings besteht dabei die Gefahr, daß man ein inhaltlich sinnloses Modell schätzt. Übrigens kann das Problem der Nichtexistenz von Maximum-Likelihood-Schätzern nicht nur bei der Analyse von Kontingenztafeln, sondern auch bei der Analyse von Individualdaten mit metrischen und/oder nicht-metrischen Regressoren auftreten. Eine detaillierte Analyse der statistischen Implikationen von strukturellen Nullen und Stichprobennullen findet der Leser bei Albert und Anderson (1984).

## 4.2 Modellspezifikationsprobleme

Die Auswahl einer geeigneten Modellklasse basiert in der Regel auf theoretischen a priori Entscheidungen, bei denen etwa die Skalenniveaus der abhängigen und der unabhängigen Variablen neben anderen Daten- und Stichprobeneigenschaften einbezogen werden. Ist ein geeignetes Modell oder eine geeignete Modellklasse ausgewählt worden, so verbleibt für den Anwender das Problem, eine Auswahl zwischen geeigneten Regressoren und Designmatrizenformulierungen (= Parametrisierung von Haupt- und Interaktionseffekten) zu treffen. Für diesen Zweck stehen drei unterschiedliche, aber eng verwandte Modellspezifikationsprozeduren zur Verfügung:

1. Zur Evaluation von Parametrisierungen werden häufig formale Signifikanztests herangezogen, mit denen man auf der Basis eines a priori gewählten Signifikanzniveaus *wichtige* Parameter bzw. Variablen einschließt und *unwichtige* ausschließt. Bei dieser Vorgehensweise kann eine Reihe von Problemen entstehen, obwohl diese (rein mechanische) Technik recht

einfach anzuwenden ist. Das erste Problem ist die Selektion einer geeigneten Evaluationsstrategie. Einige Statistiker propagieren die *Bottom-UP-Technik*, bei der man - beginnend mit dem einfachsten Modell (dem Mittelwertmodell) - sukzessiv immer weitere Regressoren bzw. Effekte einschließt. Andere wiederum starten mit dem komplexesten Modell (einem saturierten Modell) und eliminieren sequentiell insignifikante Parameter. Obwohl beide Evaluationstrategien Vorteile bieten, steht jeder Anwender vor dem Problem, sich für eine der beiden Strategien zu entscheiden, die übrigens fast immer zu unterschiedlichen Ergebnissen führen. Ein weiteres Problem bei der Anwendung von Signifikanztests ist das multiple Testproblem (Savin 1984). Theoretisch geht man bei der Herleitung von Signifikanztests davon aus, daß lediglich *ein einziger* Test an einem Datensatz durchgeführt wird. Dies entspricht einer rein konfirmatorischen Fragestellung. Bei explorativen Untersuchungen muß man Tests jedoch in der Regel wiederholt anwenden. Daher muß man im Prinzip multiple Testgrenzen auf der Basis der Bonferroni-Ungleichung oder aufgrund ähnlicher Wahrscheinlichkeitsabschätzungen verwenden. Außerdem muß sich ein Anwender für ein adäquates Signifikanzniveau entscheiden. Diese Auswahl ist recht problematisch, da a priori keine hinreichenden Gründe zur Festsetzung von Signifikanzniveaus zur Verfügung stehen.

2. Als weiteres Hilfsmittel zur Modellselektion werden häufig Determinationsmaße herangezogen, die in Analogie zum Bestimmtheitsmaß  $R^2$  der klassischen Regressionsanalyse hergeleitet werden. Klassisches Problem dieser Determinationsmaße ist, daß der Einbezug zusätzlicher Parameter grundsätzlich zu einem Anstieg des Maßes führt. Daher muß man ebenso wie bei der Anwendung von Signifikanztest Diskriminationsgrenzen verwenden, deren Festlegung ein nichttriviales Problem darstellt. Die geeignetste Verwendung von Determinationsmaßen ist der Vergleich von nicht-genesteten (nicht-geschachtelten) Modellen, bei denen man mit Ausnahme von Cox-Tests keine Signifikanztests heranziehen kann.

3. Zur Aufdeckung der Effekte von Ausreißern, Nichtlinearitäten und unberücksichtigten Variablen kann man häufig eine graphische Repräsentation von Residuen verwenden. Allerdings erfordert die Inspektion von Residuen einige Erfahrung in der empirischen Datenanalyse. Außerdem müssen Residuen in Abhängigkeit vom ausgewählten Modell geeignet skaliert werden. Daher steht kein eindeutig definiertes Residuum für alle Modelle zur Verfügung, das für alle Modelle analog zu interpretierende Aussagen liefert.

Allgemeine Strategien zur Selektion von Submodellen findet man in McCullagh und Nelder (1983) sowie in Arminger und Küsters (1986), in denen auch weitere Literaturhinweise angegeben sind.

### 4.3 Softwarepakete zur Analyse von qualitativen Variablen

Mittlerweise enthalten fast alle statistischen Softwarepakete einen oder mehrere Schätzer und Methoden zur Analyse von nichtmetrischen abhängigen Variablen. Einen detaillierten Vergleich von Softwarepaketen zum Stand 1985 findet der Leser in Arminger und Küsters (1986). Die Auswahl eines Systems hängt im wesentlichen von den individuellen Bedürfnissen eines Nutzers in Hinblick auf Kriterien wie Problemgröße, Parametrisierungsoptionen und Programmflexibilität ab. Großrechner-Systeme sind in der Regel in Bezug auf die Problemgröße nahezu unbeschränkt, während PC-Programme tendenziell einfacher zu benutzen sind. Daher beschränken sich die folgenden Bemerkungen auf Kerneigenschaften von ausgewählten und teilweise weitverbreiteten Systemen:

1. GLIM (Payne 1985, McCullagh und Nelder 1983 ) kann zur Schätzung von binären Modellen wie das Logit- und das Probitmodell, aber auch zur Schätzung von komplementären Log-Log-Modellen herangezogen werden. Designmatrizen zur Parametrisierung von nominalskalierten unabhängigen Variablen mit Eckpunkteffekten kann man mit GLIM durch eine sehr einfach handhabbare FIT-Direktive spezifizieren. Zum Aufbau der Haupt- und Interaktionseffekte für zwei kategoriale unabhängige Variablen  $A$  und  $B$  muß man lediglich den Befehl  $A * B$  (bzw.  $A + B + A.B$ ) angeben, durch den die Designmatrix automatisch aufgebaut wird. Außerdem kann man mit GLIM Maximum-Likelihood-Schätzer für loglineare Modelle zur Analyse von mehrdimensionalen Kontingenztabellen berechnen, so daß man die Parameter von multinomialen Logitmodellen indirekt durch die Spezifikation von äquivalenten loglinearen Modellen bestimmen kann, sofern alle Variablen nominalskaliert sind. Das Programmpaket GLIM steht für eine sehr breite Reihe von Maschinen zur Verfügung, zu denen unter anderem Großrechner von IBM, DEC und CDC sowie PC's gehören.
2. LIMDEP (Greene 1986) kann fast alle in diesem Artikel erwähnten Modelle schätzen. Die einzige Ausnahme ist die Schätzung des multinomialen Probitmodells, das allerdings auch mit keinem anderen Standard-

paket geschätzt werden kann. Trotz dieser breiten Spannweite ist es mit LIMDEP relativ zeitraubend, kategoriale unabhängige Variablen in die Analysen einzubeziehen, da Designmatrizen nicht automatisch aufgebaut werden können. Anstelle dessen muß man alle Designmatrizen mit Hilfe der Matrix-Makro-Kommandos in LIMDEP selbst aufbauen, was bei komplexen Designs recht mühsam ist. Innerhalb von LIMDEP wird als Schätzverfahren durchweg Maximum-Likelihood verwendet. LIMDEP steht unter anderem für Großrechner von IBM und DEC sowie für PC's zur Verfügung.

3. SPSS enthält eine Option zur Schätzung von loglinearen Modellen mit zentrierten Effekten. Dieser Ansatz ist recht nützlich, wenn man eine Kontingenztafel ohne Stichprobennullen und strukturelle Nullen analysieren möchte.

4. SAS enthält eine Reihe von Prozeduren zur Schätzung von binären und ordinalen Modellen sowie von loglinearen Modellen und multinomialen Logitmodellen. Designmatrizen können mit einem ähnlichen Syntax wie in GLIM aufgebaut werden. Allerdings werden zentrierte Reparametrisierungen verwendet, so daß man Probleme mit Nullbesetzungen bekommen kann. SAS verwendet eine automatische Aggregation der Daten, die zu Problemen führen kann (Küsters 1988b). Im Gegensatz zu den anderen Paketen ermöglicht SAS die recht einfache Durchführung von Tests auf marginale Homogenität in Kontingenztafeln. Aufgrund der in SAS enthaltenen Matrixsprache kann man allerdings alle in diesem Artikel erwähnten Schätzer selbst implementieren, sofern man einige grundlegende numerische Verfahren kennt.

5. BMDP enthält die Optionen zur Schätzung von loglinearen Modellen und binären Modellen. Unter anderem enthält BMDP auch eine Routine zur automatischen Erzeugung von zentrierten Designmatrizen.

Multinomiale Probitmodelle können mit keinem der erwähnten Programmsysteme geschätzt werden, da der numerische Aufwand zu hoch ist. Mit Ausnahme dieses Modells kann man daher lediglich mit LIMDEP alle in diesem Artikel behandelten Modelle schätzen. Aufgrund der Probleme bei der Erstellung von Designmatrizen ist GLIM jedoch bedeutend benutzerfreundlicher, wenn alle Variablen nominalskaliert sind, so daß man ein loglineares Modell zur Schätzung heranziehen kann.

Eine weitere Alternative zur Schätzung von Maximum-Likelihood-Schätzern besteht in der Verwendung einer allgemeinen Matrixsprache wie GAUSS (Edlefsen und Jones 1988), die eine Reihe von Bibliotheksmodulen

enthält, die u. a. auch binäre und ordinale Probit- und Logitmodelle sowie loglineare Modelle und multinomiale Logitmodelle umfaßt. GAUSS ist einfach zu erlernen, so daß der Benutzer die einzelnen Applikationsmodule leicht an seine Bedürfnisse anpassen kann. Die Programmierung mit GAUSS sowie eine Anwendung des binären Probitmodells findet der Leser in Küsters und Arminger (1989). Allerdings ist GAUSS nur für PC's implementiert, so daß man bei größeren Datensätzen auf andere Programmiersprachen wie etwa FORTRAN oder APL2 umsteigen muß.

## 5. Beispiel

In diesem Abschnitt wird die Schätzung eines binominalen Logitmodells anhand eines historischen Datensatzes demonstriert. Der Datensatz wurde der umfangreichen Studie über *Struktur und Handeln parlamentarischer Führungsgruppen in Deutschland und Frankreich 1848/49* von Best (1986) entnommen. Ziel der Analyse ist die Bestimmung der wichtigsten sozioökonomischen und politischen Einflußfaktoren auf das Abstimmungsverhalten der Abgeordneten der Frankfurter Nationalversammlung von 1848/49, das mit Hilfe eines hier nur kurz erläuterten Verfahrens auf eine dichotome Links-Rechts-Variable reduziert wurde. Die folgende Analyse verwendet die in Thome (1989) beschriebene vierdimensionale Kontingenztafel<sup>6</sup> und basiert auf den Datensätzen von 743 Abgeordneten<sup>7</sup>. Folgende Variablen werden verwendet:

- |      |  |
|------|--|
| KONF | Konfessionszugehörigkeit mit den Ausprägungen:<br>1 Katholisch.<br>2 Nicht katholisch (Vorwiegend Protestantisch).   |
| VK   | Verfassungskontext vor 1848 mit den Ausprägungen:<br>1 Abgeordneter repräsentierte einen Wahlkreis eines Einzelstaates, der bis mindestens 1848 als absolute Monarchie verfaßt war.<br>2 Abgeordneter repräsentierte einen Wahlkreis eines Einzelstaates, der schon vor 1848 eine Verfassung erhalten hatte. |

<sup>6</sup> Siehe Thome (1989), S. 130.

<sup>7</sup> Insgesamt waren in der Nationalversammlung 809 Abgeordnete vertreten. Abgeordnete mit fehlenden Werten wurden aus der Analyse ausgeschlossen.



- POLERF** Politische Erfahrungen des Abgeordneten vor 1848 mit den Ausprägungen:
- 1 Vor 1848 ausschließlich in politischen Ämtern tätig.
  - 2 Vor 1848 in politischen Ämtern tätig *und* an illegalen Aktivitäten beteiligt.
  - 3 Vor 1848 ausschließlich an illegalen Aktivitäten beteiligt.
  - 4 Keine politischen Erfahrungen vor 1848.
- LRO** Links-Rechts-Orientierung mit den Ausprägungen:
- 1 Rechts.
  - 2 Links.

Die Links-Rechts-Skala wurde mit Hilfe von Indikatoren konstruiert, die das Abstimmungsverhalten zu Fragen der Gültigkeit von Grundrechten, Ausschluß von Wahlrechten, Versammlungsverboten bei Gefahren für öffentliche Sicherheit und Ordnung etc. beschreiben. Die Daten sind in der folgenden Kontingenztafel, die auch eine Stichprobennull enthält, aufgelistet.

	KONF1				KONF2				
	VK1		VK2		VK1		VK2		
	LR01	LR02	LR01	LR02	LR01	LR02	LR01	LR02	
POLERF1	11	6	19	13	29	8	23	35	
POLERF2	5	1	0	8	15	3	19	25	
POLERF3	13	22	4	2	29	16	12	23	
POLERF4	103	72	26	11	94	35	23	38	

Mit Hilfe einer deskriptiven Analyse durch die bivariaten Kontingenztafeln (KONF, LRO), (VR, LRO) und (POLERF, LRO) kann man lediglich erkennen, daß der Verfassungstext VK einen erheblichen Einfluß auf die ideologische Orientierung LRO der Parlamentarier ausübt. Die folgenden Tabellen geben die konditionalen Verteilungen von LRO bei gegebenen unabhängigen Variablen an:

+-----+-----+-----+				+-----+-----+-----+			
		LR01	LR02			LR01	LR02
+-----+-----+-----+				+-----+-----+-----+			
	KONF1	0.5728	0.4272		VK1	0.6472	0.3528
+-----+-----+-----+				+-----+-----+-----+			
	KONF2	0.5714	0.4286		VK2	0.4484	0.5516
+-----+-----+-----+				+-----+-----+-----+			

+-----+-----+-----+				
		LR01	LR02	
+-----+-----+-----+				
	POLERF1	0.5694	0.4306	
+-----+-----+-----+				
	POLERF2	0.5132	0.4868	
+-----+-----+-----+				
	POLERF3	0.4793	0.5207	
+-----+-----+-----+				
	POLERF4	0.6119	0.3881	
+-----+-----+-----+				

Die konditionalen Verteilungen für LRO stimmen somit für katholische und nichtkatholische Abgeordnete fast überein. Daher stützt diese Tabelle zunächst die These, daß die Konfessionszugehörigkeit keinen Einfluß auf die politische Orientierung aufwies. Die folgende dreidimensionale Tabelle, in der die konditionalen Verteilungen von LRO für alle Kombinationen zwischen Verfassungskontext und Konfession angegeben wird, zeigt hingegen, daß diese Hypothese nicht korrekt ist.

+-----+-----+-----+					
	KONF1		KONF2		
+-----+-----+-----+					
	LR01	LR02	LR01	LR02	
+-----+-----+-----+					
	VK1	0.5665	0.4335	0.7293	0.2707
	VK2	0.5904	0.4096	0.3889	0.6111
+-----+-----+-----+					

Die auf VK und KONF konditionierte Verteilung der Links-Rechts-Orientierung LRO ist gegenüber der konditionierenden Variablen VK nahezu invariant. Somit reagieren katholische Abgeordnete offenbar nur in einem sehr geringen Ausmaß auf den Verfassungskontext. Bei nichtkatholischen Abgeordneten (vorwiegend Protestanten) ist der Verfassungskontext hingegen ein dominanter Faktor für die politische Orientierung. Möglicherweise nahmen

nichtkatholische Abgeordnete ihre Umgebung stärker als katholische Abgeordnete wahr.

Die bivariate Kontingenztafel (POLERF, LRO) zeigt einen Zusammenhang zwischen der politischen Erfahrung und der Rechts-Links-Orientierung. Abgeordnete ohne politische Erfahrung neigten eher zu einem rechtsorientierten Abstimmungsverhalten, während Abgeordnete mit einer subversiven Vergangenheit eher eine linke Orientierung aufwiesen. Die Stärke dieses Effektes muß allerdings durch ein multivariates Modell beurteilt werden, in dem alle unabhängigen Variablen berücksichtigt werden.

Zur Bestimmung der gemeinsamen Wirkung der Einflußfaktoren wurde eine Sequenz von binären Logitmodellen mit GLIM 3.77 geschätzt. Der folgende Ausschnitt aus der Log-Datei enthält die Schätzung des ausgewählten Modells. [i] und [o] sind Prefixmarker für Eingabekommandos bzw. Glim-Ausgaben. Die unabhängigen Variablen KONF, VK und POLERF wurden mit den Ausprägungsziffern kodiert, während R und L die Häufigkeiten von rechts - bzw. linksorientierten Abgeordneten für jede Kombination der unabhängigen Variablen angeben. Die Anzahl der Kombinationen wird in der UNITS Direktive festgelegt, während die Anzahl der Ausprägungen jeder unabhängigen Variablen durch die FAC Anweisung spezifiziert wird. Durch die Anweisung CAL TOTAL=R+L wird die Häufigkeit TOTAL jeder Kombination der unabhängigen Variablen berechnet, die in der Anweisungsfolge ERROR B TOTAL und YVAR L zur Spezifikation eines binären Logitmodells mit klassifizierten Daten benötigt wird. Durch die FIT Anweisung wird ein Modell geschätzt, bei dem die unabhängigen Variablen durch einen varianzanalytischen Design automatisch expandiert werden. Die abhängige Variable ist der Anteil der linken Abgeordneten innerhalb jeder Kombination der unabhängigen Variablen. Das Kommando FIT KONF\*VK+POLERF bewirkt die Schätzung eines Modells, das alle Haupteffekte von KONF, VK und POLERF enthält und in dem zusätzlich alle Interaktionseffekte zwischen KONF und VK berücksichtigt werden. Durch die DISPLAY Anweisung können verschiedene Schätzer und Statistiken selektiv ausgegeben werden.

```

[i] $UNITS 16
[i] $FAC KONF 2 VK 2 POLERF 4
[i] $DATA KONF VK POLERF R L
[i] $READ
[i] 1 1 1 11 6
[i] 1 1 2 5 1
[i] 1 1 3 13 22
[i] 1 1 4 103 72
[i] 1 2 1 19 13
[i] 1 2 2 0 8
[i] 1 2 3 4 2
[i] 1 2 4 26 11
[i] 2 1 1 29 8
[i] 2 1 2 15 3
[i] 2 1 3 29 16
[i] 2 1 4 94 35
[i] 2 2 1 23 35
[i] 2 2 2 19 25
[i] 2 2 3 12 23
[i] 2 2 4 23 38
[i] $CAL TOTAL=R+L
[i] $ERROR B TOTAL
[i] $YVAR L
    >>> Schaetzung von verschiedenen Zwischenmodellen <<<
[i] $COM AUSGEWAELHTES MODELL
[i] $CAL P3=%IF(%EQ(POLERF,3),1,0)$
[i] $COM KONF*VK+P3 --> LRO
[i] $FIT KONF*VK+P3$
[o] scaled deviance = 22.396 at cycle 3
[o] d.f. = 11
[i] $DISPLAY ME
[o] Current model:
[o] number of units is 16
[o] y-variate L

```

```

[o] weight      *
[o] offset      *
[o] probability distribution is BINOMIAL
[o]   with binomial denominator TOTA
[o]           link function is LOGIT
[o]           scale parameter is 1.000
[o] terms = 1 + KONF + VK + P3 + KONF.VK
[o]           estimate      s.e.      parameter
[o]   1      -0.3465      0.1366      1
[o]   2      -0.7550      0.2005      KONF(2)
[o]   3      -0.05727     0.2605      VK(2)
[o]   4       0.5161      0.2094      P3
[o]   5       1.525       0.3337      KONF(2).VK(2)
[o] scale parameter taken as 1.000

```

Das ausgewählte Modell  $KONF*VK + (POLERF=3)$  zeigt einige interessante Effekte, die man teilweise auch schon bei der Analyse der bi- und tri-variaten Tabellen vermuten konnte. Der Parameter mit dem Namen 1 und dem Schätzwert  $-0.3465$  repräsentiert die Neigung zu linken Abstimmungspositionen innerhalb der Referenzkategorie, die durch die Kombination der unabhängigen Variablen  $KONF=1$ ,  $VK=1$  und  $POLERF=1$  gegeben ist. Das negative Vorzeichen bedeutet, daß diejenigen katholischen Abgeordneten aus absolutistischen Staaten, deren politische Erfahrungen vor 1848 sich auf die Ausübung offizieller Ämter beschränkte, wahrheitlich nach rechts tendierten. Der Parameter  $KONF(2)$  mit dem Schätzwert  $-0.7550$  gibt die Verschiebung der Kombination  $KONF=2$ ,  $VK=1$  und  $POLERF=1$  gegenüber der Referenzkategorie an. Dies impliziert, daß nichtkatholische Abgeordnete gegenüber katholischen Abgeordneten innerhalb der Gruppe der Parlamentarier mit monarchistischem Verfassungskontext und politischen Erfahrungen durch politische Ämter eine noch höhere Neigung zu rechten Abstimmungspositionen aufwiesen. Der nichtsignifikante Parameterschätzer für  $VK(2)$  wird lediglich benötigt, um den signifikanten Interaktionseffekt  $KONF(2) \cdot VK(2)$  zu schätzen. Ein Einfluß des Verfassungskontexts innerhalb der Gruppe der Parlamentarier mit katholischer Konfession auf die Links-Rechts-Orientierung ist hingegen nicht feststellbar. Innerhalb der Gruppe der Parlamentarier, die einen Wahlkreis repräsentierten, der vor 1848 eine Verfassung erhielt, ist hingegen ein interessanter Effekt erkennbar. Nichtkatholische Abgeordnete neigten innerhalb dieser Gruppe

eher zu linken Abstimmungspositionen, wie man aus dem positiven Parameterschätzer 1.525 für KONF (2) .VK (2) schließen kann.

Eine Auswirkung des politischen Erfahrungshorizontes zeigt sich im Datensatz in einer signifikanten Form lediglich bei denjenigen Abgeordneten, die sich vor 1848 subversiv betätigten. Abgeordneten mit subversivem Hintergrund neigten offensichtlich zu einem verstärkt linken Abstimmungsverhalten, wie der zur Dummy-Variablen P3 korrespondierende Parameterschätzer 0.5161 zeigt. Abgeordnete aus den anderen Gruppen (POLERF=1), (POLERF=4) sind hingegen relativ homogen, wenn man die Variablen KONF und VK berücksichtigt. Somit ist die aus der marginalen Kontingenztafel zwischen POLERF und LRO ablesbare Rangfolge POLERF=4, POLERF=1, POLERF=2, POLERF=4 ein Scheineffekt, der lediglich durch die Aggregation über alle Ausprägungen der Variablen KONF und VK entsteht.

Für die nicht abgedruckten Modelle wurden folgende Devianzstatistiken berechnet. Differenzen von Devianzstatistiken zwischen genesteten Modellen sind approximativ  $\chi^2$  verteilt, wobei die Anzahl der Freiheitsgrade durch die Differenz der Freiheitsgrade zwischen den konkurrierenden Modellen bestimmt wird.

Modell	Devianz	df (Freiheitsgrade)
1	79.744	15
KONF	79.742	14
VK	51.574	14
POLERF	71.839	12
KONF+VK	50.369	13
KONF*VK	28.493	12
VK+POLERF	44.756	11
VK*POLERF	37.052	8
KONF+VK+POLERF	43.010	10
KONF*VK+POLERF	22.296	9
KONF*VK+(POLERF=3)	22.396	11

Die Insignifikanz zwischen den Kategorien POLERF=1, POLERF=2 und POLERF=4 ist auch aus der Devianzdifferenz zwischen den Modellen KONF\*VK+POLERF (3) und KONF\*VK+POLERF ersichtlich, die bei 2 Freiheitsgraden 0.1 beträgt und somit nicht signifikant ist. Übrigens lassen sich die Parameter im saturierten Modell KONF\*VK\*POLERF nicht bestimmen, da die Stichprobennull in der Kontingenztafel zu einer Nichtexistenz des ML-Schätzers im saturierten Modell führt.

## 6. Schlußfolgerungen

Regressionsmodelle zur Analyse von nichtmetrischen abhängigen Variablen können beim gegenwärtigen Stand der statistischen Modellierung ohne erhebliche Probleme angewendet werden. Allerdings sind Schätzer für die meisten Modelle bis zum jetzigen Zeitpunkt noch nicht in einer benutzerfreundlichen Form implementiert. Dies erzwingt beim Anwender eine Auseinandersetzung mit den typischen Analyseproblemen wie Stichprobennullen etc., da automatische Hinweise auf Datenprobleme von gegenwärtigen Softwareprodukten noch nicht gegeben werden.

## Literatur

- Albert, A. und Anderson, J.A. (1984): On the Existence of Maximum Likelihood Estimates in Logistic Regression Models. *Biometrika* 71, 1 - 10.
- Amemiya, T. (1973): Regression Analysis when the Dependent Variable is Truncated Normal. *Econometrica* 41, 997- 1016.
- Amemiya, T. (1985): *Advanced Econometrics*. Oxford.
- Anderson, J.A. (1984): Regression and Ordered Categorical Variables (with discussion). *Journal of the Royal Statistical Society B*, 1 -30.
- Arminger, G. und Küsters, U. (1986): Statistische Verfahren zur Analyse qualitativer Variablen. *Forschungsberichte der Bundesanstalt für Straßenwesen* 147. Aachen.
- Best, H. (1990): *Die Männer von Bildung und Besitz. Struktur und Handeln parlamentarischer Führungsgruppen in Deutschland und Frankreich 1848/49*. Bd. 90 der Schriften zur Geschichte des Parlamentarismus und der politischen Parteien. Düsseldorf.
- Bishop, Y.M.M., Fienberg, S.E. und Holland, P.W. (1975): *Discrete Multivariate Analysis - Theory and Practice*. Cambridge.
- Bock, R.D. (1975): *Multivariate Statistical Methods in Behavioral Research*. New York.
- Daganzo, C. (1979): *Multinomial Probit - The Theory and Its Application to Demand Forecasting*. New York.

- Edlefsen, L. und Jones, S. (1986): GAUSS-Programming Language Manual. Kent, Aptech Systems Inc.
- Fahrmeir, L. und Hamerle, A. (1984, Hrsg.): Multivariate statistische Verfahren. Berlin.
- Greene, W. (1986): LIMDEP-Reference-Manual. New York University.
- Jarausch, K.H. und Arminger, G. (1988): The German Teaching Profession and Nazi Party Membership: A Demographic Logit Model. *Journal of Interdisciplinary History*, 1989, 197-225.
- Jarausch, K.H., Arminger, G. und Thaller, M. (1985): Quantitative Methoden in der Geschichtswissenschaft - Eine Einführung in die Forschung, Datenverarbeitung und Statistik. Darmstadt.
- Küsters, U. (1988a): LIMDEP - Ein vielseitiges Programmsystem zur Schätzung von Regressionsmodellen für abhängige Variablen beliebigen Meßniveaus. In: Faulbaum, F. und Uehlinger, H.M.: Fortschritte der Statistik-Software 1. Stuttgart.
- Küsters, U. (1988b): Analyse von qualitativen Variablen. In: Faulbaum, F.: Datenanalyse mit SAS. Erscheint beim Gustav Fischer, Stuttgart.
- Küsters, U. und Arminger, G. (1989): Programmieren in GAUSS - Eine Einführung in das Programmieren statistischer und numerischer Algorithmen. Stuttgart.
- Maddala, G.S. (1983): Limited-dependent and Qualitative Variables in Econometrics. Cambridge.
- McCullagh, P. (1980): Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society B20*, 109- 142.
- McCullagh, P. und Nelder, J.A. (1983): Generalized Linear Models. London.
- McFadden, D. (1974): Conditional Logit Analysis of Qualitative Choice Behavior. In: Zarembka, P. (ed.): *Frontiers in Econometrics*. New York, 105-142.
- McFadden, D. (1981): Econometric Models of Probabilistic Choice. In: Manski, C.F. und McFadden, D. (Hrsg.): *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge, Massachusetts, 198-272.
- Payne, C.D. (1985, Ed.): The GLIM System Release 3.77. Oxford, NAG-Royal Statistical Society.
- Röding, M., Küsters, U. und Arminger, G. (1985): KALOS-Programmbeschreibung - Ein interaktives Programmsystem zur Analyse kategorialer Logitmodelle. Köln, Zentralarchiv für empirische Sozialforschung.
- Savin, N.E. (1984). Multiple Hypothesis Testing. In: Griliches, Z. und Intriligator, M.D.: *Handbook of Econometrics - Volume II*. Amsterdam, 827-879.



- Schmidt, P. (1976). *Econometrics*. New York.
- Stewart, M.B. (1983): On the least squares estimation when the dependent variable is grouped. *Review of Economic Studies* L, 737 -753.
- Thisted, R.A. (1988): *Elements of Statistical Computing*. New York.
- Thome, H. (1989): *Grundkurs Statistik für Historiker, Teil I: Deskriptive Statistik*. *Historische Sozialforschung*, Beiheft No. 2., 1 - 147.
- Tobin, J. (1958): Estimation of relationships for limited dependent variables. *Econometrica* 26, 24-36.